

ABSTRACT OF THE DISCLOSURE

A parsing system and method are provided in which the break characters in the document are used to rapidly parse the document and extract one or more key phrases from the document which characterize the document. The break characters in the document may include explicit break characters, such as punctuation, soft stop words and hard stop words. The determination of which phrases in the document are extracted depends upon the type of break character appearing after the phrase in the document. The parser may also be used to parse a foreign language document into one or more phrases.

DOCTEST-100